

A

F

T

12 Sep 67

HISTORY OF PROJECT CHIVE

A. Before Phase I

The rationale for the CHIVE Project had its roots in numerous investigations, conducted by in-house as well as contractor personnel since approximately 1959, all of which generally agreed both as to the increasing seriousness of the information retrieval problem within the Agency and the advisability of introducing new hardware and techniques reflecting the present state-of-the-art of the information-handling technology. In a report issued in 1958 under the auspices of the Central Reference Advisory Group (CRAG), it was stated that in one DD/I Office (OCR):

"The volume of incoming information exceeds processing capabilities based on current manual or electrical accounting machine (EAM) techniques;

The proportion of receipts which can be fully processed is declining;

Service from existing facilities is becoming slower as the size of the several indexes increases;

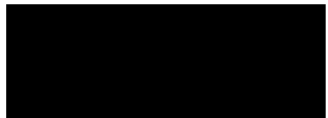
Quality of service in terms of listing, subject correlation, up-dating, and display is declining or not offered because of the limitations of current staff and equipment."

These findings with respect to OCR were complemented by indicated applications for computers in ORR, OSI, and other DD/I Offices.

The CHIVE Project ~~to date~~, therefore, was not envisioned as a study of the applicability of computers and other devices to the document handling problems of the Agency, but rather as a system-oriented attempt to reduce the spectrum of DD/I problems to manageable proportions in order to exact from available equipment and techniques their known design potential in the solution of these critical problems. Because the magnitude and complexity of the task equaled or exceeded any previous design effort in the storage and retrieval area, the search for problem solutions was characterized by an attitude of relative caution.

The major events preceding the initiation of Phase I of the CHIVE project are as follows:

1. Feb 1959 CIA Management Staff recommended to the DCI that a study be undertaken of the computer requirements of the DD/I area.
2. Jul - Aug 1959 DD/I components conducted a pre-feasibility study, concluding that there are areas demanding improvement for which ADP offers significant potential support, and recommending "that an EDPM feasibility study for the DD/I area be undertaken."
3. Jan - Mar 1961 [REDACTED] under the sponsorship STATINTL
of the DD/I Automation Staff, conducted 6-week surveys of the DD/I, finding that the DD/I was seriously behind other Community members in the information handling art and recommending that a large-scale effort be launched immediately with heavy contractor participation.

- 
4. April 1961 The DD/I Automation Staff, issued a paper recommending that the DD/I establish a Computer Center, implement production applications on the computer in specific problem areas, and undertake a major system study of the overall document and information retrieval problem.
 5. July 1961 The CIA Automatic Data Processing Staff (ADPS) was created in the DD/S and the DD/I Automation and CIA Management Staffs abolished.
 6. October 1961 Formal management review was completed and approval received for an EDP development program for the DD/I based upon the Automation Staff's proposals of April 1961.
 7. January 1962 ADPS published a document describing its planned development program for the DD/I in greater detail. The principal phases of the system design effort in the document and information retrieval field (Task III of the proposed program) were anticipated to include: (a) contractor orientation, (b) planning the study, (c) fact-finding, (d) system design, (e) testing of techniques and models, (f) implementation of selected sub-systems, and (g) documentation.

It should be emphasized here that the name CHIVE was originally applied to a three-part program aimed at moving ^{the Intelligence Directorate} CIA into the computer environment.

- (1) The first element of the program was to establish a computer center as such, and to acquire the personnel, hardware, and expertise necessary to operate this center.

- 4 -

(2) The second element was to implement specific ADP applications which had been identified by study groups over a period of a year or so. (Note: The intent of these first two elements was to get started on specific, well-defined applications as soon as ADP support was available.)

(3) The third element was to examine the general information/document retrieval problem in the Agency.

The goal of this third element was later altered to address the information/document retrieval problems of the DD/I and particularly the Office of Central Reference (now known as the Central Reference Service). Thus modified, the third element of the three-part program became Project CHIVE

It was apparent to CHIVE architects that though the Agency had senior and capable people who were intimately aware of the capabilities and shortcomings of its present information storage and retrieval system, it was woefully poor in ADP talent—both system analysts and programmers. This was evidenced by the willingness of the Agency to ask contractors such as [REDACTED] acknowledged^{STATINTL} as proficient in system design and development, to study the Agency's problem.



- 6 -

B. Phase I

The first phase of a multi-phase system development was a requirements determination phase and fact-finding survey. Approximately 150 interviews of user offices--primarily research or production offices--in CIA were conducted during this period. These were aimed at determining what information problems existed, how the user regarded the present central reference system, what source material were of primary concern, and in general what requirements would be levied on system developers. One of the principal issues was the existence of a central reference facility. Alternative methods of providing analyst support existed and had points in their favor. Separate and distinct special project files was one alternative. Another was to provide assistance to the research and production analysts in maintaining their own personal collections--the so-called "shoebox files".

The analysts conducting and evaluating the interviews had to be as objective as possible since the identification of which kind of system users preferred was a major influence in establishing requirements. It turned out that the majority preferred the central reference facility to the other alternatives. With this basic concept determined, more detailed requirements could be set forth.

The major events of Phase I were:

May - Jul 1962	Clearance and orientation of contractor personnel, as well as recruitment and preliminary training of initial "hard core" element of ADP Staff, was completed.
----------------	--

- 7 -

Aug 1962	The scope and methodology of the fact-finding effort was developed and management approval obtained for a detailed survey of all DD/I offices.	
Sep - April 1963	Joint ADPS/ [REDACTED] teams interviewed approximately 150 component organizations of the DD/I.	STATINTL
Oct 1962	A memo from Ray Cline (DD/I) to Intelligence Assistant Directors reviewed the three CHIVE task areas and endorsed the project efforts.	
Jun 1963	Major findings and recommended design parameters resulting from the Phase I DD/I system study were presented to selected DD/I managers in oral and written form.	
Aug 1963	The DD/I formed a CHIVE Evaluation Group, composed of representatives from various Agency offices, to review the CHIVE Phase I report.	
Aug 1963	The Automatic Data Processing Staff was transferred from the DD/S to the newly established DD/S&T and reorganized as the Office of Computer Services.	
Sep - Oct 1963	The CHIVE Evaluation Group completed its analysis of the CHIVE Phase I findings, recommending (<u>inter alia</u>) that the Project continue into a second phase with guidance and support from the DD/I and careful testing of system feasibility prior to implementation.	

The report entitled, Project CHIVE--Preliminary System Design Report--Information Processing, 1 Dec 1963, produced after Phase I ended, was devoted for the most part to a discussion of System Objectives and System Functions Preliminary Specifications.

- 8 -

The two principal system characteristics were:

- (1) Document-based
- (2) Centralized control

The first of these had been denied as a necessary attribute by some staff employees with considerable experience in the Agency but who were ambitious to push forward to a system of data control, data manipulation, and data on information retrieval. In effect, the proponents of this concept suggested that once the data and information content of a document had been identified, extracted, indexed, coded, and stored in an EDP system, the document was of only marginal importance. Reliance would be placed on the EDP system to retrieve by subject, category, named object, and so on the gist of the documents indexed.

Those who supported the need to build into the system a capability for document as well as fact retrieval were simply responding to the evidence supplied by a central reference service. This experience showed that analysts very often wished to read the parent document from which data had been extracted. That was viewed as a requirement which would continue for some time to come. Hence the system would be document-based.

Discussion of the second one addressed the multiplicity of repositories which characterized the present system and went on to propose that "...the specialized register concept be abandoned and the personnel of the central system be organized along geographic lines to the extent possible." Clearly, reorganization of OCR was basic and fundamental to the CHIVE design.

- 9 -

Under System Service in the same report the CHIVE designers listed the following as basic to the design: "(1) that the CHIVE system and the analyst files will complement one another, and (2) that the system's primary customer is the Agency desk analyst." Thus, the continuance of the "shoe-box file" was acknowledged together with an emphasis on user, not operator, satisfaction.

The chapter on System Scope in the Phase I report had five sections which pretty well established system boundaries. Listed below are the section headings with short excerpts from each:

"1. Document Boundaries

...The data base for the CHIVE System must, therefore, encompass all documents in use by the analytic offices which the system must serve.

2. Indexing Boundaries

...In addition to retrieving material based on multiple criteria, (i.e., personalities, organizations/installations, geographic location, subject) a capability for large-scale manipulation of the data base index must be provided to permit information to be correlated for the analyst. This, indeed, is one of the primary justifications for acquiring a computer which, unlike simpler machine tools, has both the speed and logic capabilities to perform this function effectively.

3. Security Boundaries

If the proposed system is to serve in fact as the central repository of positive intelligence information for all components of the Agency, it must process and store any information of continuing intelligence value--whatever its security classification.

4. Dissemination

The Agency must have the capability to disseminate incoming documents automatically on the basis of the document content

- 10 -

and analyst's statements of interests.

5. The Initial System

The basic objective of the initial system is to establish a small-scale mechanical structure of the eventual system in a limited, controlled environment. This system will be designed with the performance specifications of the full system in mind for application to a limited area. One of the basic premises of the initial system, however, will be that expansion in terms of added sources and increased performance will not require substantial redesign of programs, methods of operation or equipment configuration."

Clearly, these concepts were controversial to many. Note in 2. the reference to the justification for a computer. In 3. there was embodied the "all-source file" concept. In 4. automatic dissemination was advocated (later abandoned in the design of a first system). In 5. the bread-board approach was recommended. Coupled with the proposed area reorganization this indicates that the initial system would be tested in one geographic area component. As the culmination of the Phase I data gathering and conceptual design effort, this document is noteworthy.

STATINTL The year 1963 was filled with accomplishments and changes for CHIVE. Two computers (1401/1410) for the DD/I were delivered in January 1963 and in subsequent monthly reports were referred to as the CHIVE Computer System. CHIVE programmers (trainees) worked under the tutelage [REDACTED] Computer support prior to the installations of the 1401/1410 Systems was provided by DDP. It is noteworthy that during the first half of 1963, Project CHIVE had the three-pronged emphasis noted earlier. In fact, a seven-page report written for the DD/I at the end of FY-63 devoted six

- 11 -

pages to accomplishments in: establishing a computer center and implementing applications. The seventh page announced the end of Phase I of the Document/Information Retrieval Systems Design--Task III - and commented that this task was the most difficult of all CHIVE tasks.

The CHIVE Evaluation Group was also constituted in July/August, and the August status report indicated that further progress on Task III - System Design had virtually stopped. In September those CHIVE personnel assigned to Task III were busy preparing briefings and reports for the CEG.

The CEG report of 7 October 1963 recommended the continuation of project CHIVE--but not without reservations. The group recommended that management responsibility for CHIVE should be vested in the Intelligence Directorate, and that OCR should continue to improve its present system. The subject of hardware received attention as well. Clearly the CEG was not concerned that program testing be done in a centralized facility such as OCS, but when computer utilization reached one shift or more, the computer system should be part of the central reference system of the Intelligence Directorate. Moreover, CHIVE was charged to consider suitability for the Intelligence Directorate of extant systems when assessing equipment requirements.

Another area of concern to the CEG was that of testing. The group seemed to be expressing misgivings about OCR plunging into a new system without testing of pilot projects. Recommendations were made to the effect that formulation of such pilot projects

- 12 -

be approved by the DD/I. (As the indexing experiment of Phase II showed, such activities are expensive, time consuming, and in the final analysis, suggest, but do not prove, an operational approach).

C. Phase II

Management review by the CHIVE Evaluation Group (Borel and

STATINTL

STATINTL

[REDACTED] of the Phase I report, the reorganization which resulted in OCS, and the physical relocation of people and equipment were major factors in causing Phase II to have a slow start. A progress report of approximately 28 pages written 18 September 1963 was a distillation of design concepts garnered in Phase I and polished in the first months of Phase II.

CHIVE 1-63, 1 Dec 1963, presented the same schedule shown in the June 1963 Project CHIVE Quarterly Report. This schedule called for the completion of Phase I (of Task III) in June 1963. The same schedule shows Phase II as running from July 1963 through June 1964 (actually the Phase II report was not completed until the Spring of 1965). Phase III in the old schedule was to run from July 1964 through April 1965--in retrospect, an incredibly optimistic plan. In fact, Phase II was completed at about the same time that Phase III had been planned to end.

The establishment of OCS and the designation of Task III as Project CHIVE resulted in a much smaller team. (It should be noted that Project CHIVE went from about 60 people to six and from three tasks to one without a name change).

This team was composed of analysts from several components. Like most projects, CHIVE was an organizational anomaly. [REDACTED] OCR, and OCS contributed manpower to the effort. This was advantageous in that it brought together the right mix of disciplines and experience. However, management guidance and control also came from various sources. This

STATINTL

was not as beneficial.

STATINTL The OCS contingent was the fledgling Development Branch (later Division); the OCR contingent, designated the CHIVE Support Staff, consisted of senior analysts who for the most part had acquired their experience in the Library (Documents Division), Biographics Register and Special Register. [REDACTED] -- the largest of the three -- consisted of analysts and programmers.

STATINTL In October 1963, project personnel had been assigned to major design areas. [REDACTED] had sole responsibility for all facets of design with the exception of Information Processing -- indexing procedures, indexing language, definition of files, document processing, file conversion, selection criteria and so on. These topics were addressed by a group of agency employees. It was recognized rather early that this was the bellwether of the design areas--decisions made here would have a marked influence on all other areas.

STATINTL [REDACTED] had submitted a contract proposal in June of 1963. Discussions between agency designers led by [REDACTED] on how to get the job done continued for some time. Clearly [REDACTED] felt that they could not proceed intelligently to work on their assigned design tasks without input from the Information Processing Task Team. Since [REDACTED] they felt they possessed some pertinent experience they wished to be represented on that team as well. The staff personnel were loath to surrender, even partially, design tasks in the one area they understood thoroughly and where they felt their contribution would be more valuable than any work done by an outsider. Sharing the burden was not particularly palatable to [REDACTED] since they preferred to work STATINTL

on well-defined tasks which were their exclusive responsibility. Also, they had a different design concept from that set forth by the in-house people. This was to be expected but was not precisely welcome.

Compromises were discussed which would permit [REDACTED] to STATINTL maintain headway in their exclusive task areas but would also let them assign personnel to the Information Processing Task. To the extent possible, this had to be reflected in the contract. The proper wording was not easy to find. As a consequence [REDACTED] contract was not signed until June 1964--one year after [REDACTED] had written their proposal. STATINTL

The August 1964 monthly report shows assignment of personnel to major design areas. Of these, [REDACTED] had sole responsibility for STATINTL EDP System Design and Programming Support. Two contractor employees were responsible for the Document Delivery System. Thus contract personnel did design work without system design responsibility. This continued to be an awkward situation from their point of view and it resulted in a fairly rapid turnover in [REDACTED] on-site managers during Phase II. STATINTL

A list of major events during Phase II is as follows:

1. Sept 1963 First progress report of Phase II produced.
2. Dec 1963 CHIVE 1-63 which was to be the design "bible" for the project was produced.
3. April 1964 [REDACTED] appointed as DD/I CHIVE Project Officer. CHIVE Evaluation Conference held at [REDACTED] resulted in substantial change to organization of system design efforts as well as decisions in design areas where alternatives had previously existed.
4. May/June 1964 First indexing experiment utilizing computer programs, conducted by CHIVE personnel.
5. June 1964 - June 65 During this period a large scale indexing experiment utilizing computer programs for

STATINTL

STATINTL

- 16 -

storage and retrieval of index records was planned, designed, staffed, conducted, and analyzed.

6. August 1964 A functional flow diagram was produced showing input/outputs and identifying program elements to be written in support of the OCR/CHIVE Indexing Experiment.
7. September 1964 A technical report entitled "A Comparative Analysis of Document Delivery Systems for Large and Active Files" was distributed, thus signalling the completion of the major design effort in this area.
8. November 1964 During the third week, indexing began on the OCR/CHIVE Indexing Experiment. The team, consisting of 26 people, had indexed 1500 documents by the end of the month.
9. December 1964 The major components of the computer program system necessary to support the Indexing Experiment were operational and the building of tape files containing index data was begun.
10. January 1965 Indexing of the 5,130 documents for the Indexing Experiment was completed and the majority of the indexers returned to their home components in OCR. Approximately nine members of the team were retained to assist in the query phase of the experiment.
11. March 1965 A preliminary study of input transcription techniques submitted by SRI for the Phase II report recommended the use of a page reader. A study of available optical character recognition systems was begun. A revision of the Intelligence Subject Code (ISC) was begun as a result of the indexing experiment. This was a Phase II task started during Phase II.
12. April 1965 The Systems Analysis Staff was formed in OCR and the CHIVE Support Staff was abolished.
13. May 1965 By the end of May, three of the seven volumes of the CHIVE Phase II report had been printed and distributed. A project was initiated to produce a China Locations Dictionary by merging extant dictionaries. This was another task area intended for Phase III.

- 17 -

14. June 1965 The query phase of the OCR/CHIVE Indexing Experiment was completed, signalling the end of the experiment and associated analysis. Another Phase III task was initiated. This one was concerned with the feasibility of converting large OCR files for CHIVE use in the operational phase. STATINTL
15. July 1965 All seven volumes of the Phase II report were disseminated. The CHIVE Project was reorganized as the CHIVE Task Force with [REDACTED] of OCS as its director, reporting to the director of OCR. Management approval of the Phase II report and direction to proceed with Phase III had not been received at this time. The design effort, which culminated during Phase III in production of the SKAN (Subject Keyword Announcement) publication, began this month.
16. August 1965 By the end of this month, most of the tasks planned for Phase III had been initiated and plans to recruit the personnel necessary for full staffing of the task force had been formulated.
17. September 1965 A memo to D/OCS/DDS&T from the DD/I dated 16 September 1965 stated that with certain exceptions the Phase II report and recommendations had been approved, and that the D/OCR had been directed to divert resources to staff the CHIVE Task Force. Thus Phase III began.

This list of events is only an effort to capture highlights.

Another list might be equally indicative of the problems and satisfactions of system design work.

STATINTL

One of the major design areas that was not highlighted in this list was the EDP System. A great deal of work was done by [REDACTED] in this area. They studied core allocation problems, interface with S/360 Operating System problems, file definition problems, and so on. Many reports and recommendations were produced. Volume VII of the Phase II

report which covered the EDP System design was 389 pages long-- exceeded only by Volume V, 526 pages. Unfortunately the detail of the design was based on other areas about which information was incomplete and/or constantly changing. For example, IBM's S/360 hardware/software dilemmas had not become public knowledge at that time. Equally important, detailed requirements were not obtainable from the customer, the DD/I analysts, nor from the operator, OCR, whose personnel had incomplete understanding of what was required in order that computer programs could be specified and written. Thus the task of specifying requirements fell to the system designer/developer--a very risky business.

The result of this approach was that an EDP system was designed which was generalized along almost any dimension that one could choose-- with at least one notable exception, the hardware system. The designers were assured that the ~~EDP~~^{EDP program} system would be run on a large scale system, i. e., an IBM 360/50 or larger since it was OCS's intention to go to that line of equipment, ~~and~~^{and} the Phase II recommendations called for the EDP ~~program~~^{program} system to run initially on OCS hardware with a review of the OCS computer load after two years to determine who should manage the computer element of CHIVE. However, [REDACTED] (DD/I) memo of 18 September ^{STATINTL} did not accept that recommendation and repeated the theme expressed in the CHIVE Evaluation Group report, i. e., the Intelligence Directorate should have its own computer. By omission it was agreed that testing would be done on OCS equipment so the program development could proceed. (As Phase III events will show, OCS's hardware was rejected and the CHIVE program system was circumvented.)

Very general information storage and retrieval systems are very complex and very difficult to build. However, the designers felt that the demands of fact/information retrieval/manipulation in an intelligence agency were sufficiently demanding to justify a very powerful system.

Another point of interest in Phase II was simply the historical problem of deciding when Phase II ended and Phase III began. The OCS/CHIVE Indexing Experiment spawned Phase III tasks while it was in progress. As soon as the indexing was completed and the indexer's critiques had been produced and absorbed, senior analysts on the experiment team were anxious to translate findings into indexing tools and procedures. So the modification of the ISC was begun. Also, analysis and testing of the indexing position was initiated to determine what indexing aids were appropriate. It should be remembered that a basic CHIVE concept called for a division of labor between analyst and indexer; many OCR analysts did both. Thus the indexing position and its relation to the analyst was fair game for analysis, study, and test.

A factor, which contributed to the initiation of Phase III tasks while experiment queries were still being tidied up, was geography. The indexing experiment had been conducted in a different part of the building than that occupied by the CHIVE Project. When the indexing was finished, the cadre of senior analysts involved in the experiment stayed in that area to analyze queries and discuss results. These people (all OCR personnel) were documentalists, index method experts, and librarians. It was very natural for them to apply experimental results. So Phase III had a running

start before the DD/I memo was written in September. The staff personnel necessary for some tasks were not available, however, and could not be obtained before management approval to proceed with Phase III.

Another aspect of the Indexing Experiment that warrants attention STATINTL was the programming support required. [REDACTED] took advantage of the Formatted File System developed by IBM for Fleet Intelligence Center Europe which could be run on the IBM 1410 in OCS^y ^{and} However, a query program, AD003, previously written by OCS programmers, was used to provide ~~the necessary~~ additional query capability. In addition, programs had to be written specifically for the experiment to provide input processing and listing functions. This collection of programs provided the operational capability necessary to process index records and provide responses to queries, including KWIC (Keyword In Context) listings.

Over and above this system were the statistical routines which were produced by the programmers to assist the analysts in their interpretation of the experimental data. These routines provided statistics on leader data, content data, document categories, and codes used by the indexers. It should be noted that either in a test or an operational situation, statistical analysis is a powerful tool in analyzing system strengths and weaknesses.

Finally there were computer programs necessary to provide dictionaries and authority lists. The indexers required such listings to do their job. In addition, the people (analysts) answering queries needed such lists in order to obtain some idea of what kinds of data were held in the file. It was this last area of computer programming that stimulated the dictionary building efforts over the next two years.

[Handwritten mark]

21

The design concepts postulated in the Phase I report and given a firmer emphasis in CHIVE 1-63, went through considerable examination and testing during Phase II. It is interesting to note that there was no substantial shift of design emphasis as a result of this exercise. One area which received only cursory attention before it was exempted from further analysis was the dissemination area. Two very good reasons existed for excluding it from the initial CHIVE system. First, the dissemination process had a good reputation with its customers and since it was readily isolatable from the main system, it seemed pointless to disturb it. Second, the contract work being performed for the Air Force at the Foreign Technology Division had as its first application the automation of dissemination. It was reasonable, therefore, to wait for the fruits of this labor.

Although the formidability of the security problems associated with an all-source EDP file were better understood by the end of Phase II, the project personnel viewed them as primarily technical in nature and capable of solution.

Study of the selection area during Phase II had started to address the nuts and bolts of the problem with no drastic shift from the document types identified earlier. Clearly, the amount of material to be captured was a direct function of manpower available. The results of the indexing experiment gave added importance to the role of subject indexing without reducing

21

the value of indexing named objects.

The political and inertial problems associated with a reorganization of OCR along geographic lines were recognized but postponed to a certain extent. The initial component, COG, would operate in a test mode for some period of time before seizing production responsibility from those several components in OCR which handled China related queries and reports.

The team of information analyst and indexer(s) had been tested to a degree during the indexing experiment but it was clear that the concept had to be tested in an operational environment before a decision could be made regarding its validity.

Such design concepts as single point service, improved document delivery system, and direct input of preformatted index data were still viable at the end of Phase II. The computer program system received little criticism and it was apparent that OCR as the operating officer would rely upon the experience and expertise of OCS to design a satisfactory and responsive system.

~~XX~~

- 24 -

D. Phase III

As indicated in the foregoing discussion, almost all of the tasks planned for Phase III had begun before management approval for Phase III had been received. In fact, a portion of the File Conversion task was completed in September 1965 with a preliminary report to the effect that Intellofax holdings were unlikely candidates for conversion to the CHIVE format.

One task that could not be accomplished prior to official sanction of Phase III was staffing. In order to staff, one had to interview, which had to follow the circulation of vacancy notices, which depended on an approved T/O, which depended on an organizational structure with position descriptions approved by OCR management and "wage and salary". Much staff work could be done in advance, of course, such as identifying the types of people needed and preparation of a list of candidates. However, management approval was mandatory in order to officially start the process and so the fact that recruiting for the OCR portion of the task force was able to begin in November of 1965 must be considered a major accomplishment.

The decision on the part of directorate management to deny the OCR request for additional slots to staff the development of the CHIVE system meant that OCR had to go through the agonizing process of taking slots away from its ^{production} substantive components. The approved manning level for the CHIVE Task Force was set at approximately 91 of which

- 25 -

OCR was committed to supply 54 (OCS - 21 and the remainder from contractors). Since this was another in a series of manpower cut-backs in the production units in OCR, it was not welcome.

It was also clear that some of the CHIVE concepts and recommendations were unattractive to personnel in OCR. Managers at various working levels were thus asked to surrender positions to a development effort they did not understand, and whose product they did not feel they needed. As a consequence, the Phase III effort, which was to be a detailed design and development activity, had some significant political hurdles to surmount beyond those of a strictly technical nature.

Briefly, the CHIVE Task Force was charged to produce during Phase III those tools and procedures necessary for a unit -- called the China Operations Group -- to handle all-source receipts on Communist China and respond to queries and production requests regarding Communist China levied on them by other CIA components. The China Operations Group (COG) was to be the first substantive component of an OCR organized along geographic lines. The process was viewed as incremental starting with COG. The next component would be a Far East Division with COG as a branch. After that, the other geographic divisions would be established in fairly rapid order as experience dictated.

The most difficult problem for the CHIVE Task Force during Phase III was the production of the computer program system. Some of the technical difficulties experienced in this area have been referred to previously. However, the management problems attendant to large program system development

were also severe. An established fact in the ADP world is the perennial scarcity of qualified and experienced programmers and design analysts. This scarcity is more pronounced in federal agencies than in industry because of pay scale and other inequities. Moreover, programmers with experience in designing and writing programs for information storage and retrieval applications are a rare commodity in any environment. The CHIVE programming group was therefore in a very difficult position. OCR had no experienced programmers to contribute. Some trainees were acquired from OCR, but they were unable to help design or specify programs. OCS had some experienced people, but the demands for programming support within the agency precluded detailing senior programmers to CHIVE. [REDACTED] had, or could acquire, programmers, but once more, the desire on the part of CHIVE management to maintain design control dictated a strong participation by staff employees. This left recruiting in the "market place" as a means of getting programmers. A most unattractive prospect. Recruiting actually started in Phase II and even with this advantage, Phase III began with only two staff employees in the programming unit -- the chief and one analyst/programmer. In June 1967, over a year after recruiting had begun, the contingent of staff employees in the programming unit had grown to 11 -- 4 trainees, 6 programmers or programmer/analysts, and the unit chief. The trainees, in a practical sense, were often more of a handicap than a help since their need for on-the-job training diluted the efforts of the more experienced and productive personnel.

STATINTL

- 27 -

This unit never achieved its assigned T/O manning level.

The last major blow to the programming effort occurred a short time before the task force was disbanded. (It should be noted, however, that this event was not the cause of or even a major contribution to OCR's decision to abolish the project.) The result of this disruption was a further slipping of the expected completion date of the program system and the account of this event is included here to indicate one of the disadvantages of the task force concept.

STATINTL
STATINTL
The programming group, as noted earlier, was made up of people from OCR, OCS, and [REDACTED]. One major area of development and detailed design was assigned to [REDACTED] who directed the activities of a section consisting largely of other [REDACTED] and trainees. STATINTL
As detailed program specifications were completed and coding began, it became clear to the OCS managers of the group that this unit had produced a design which was unresponsive to the total system requirement. After considerable study and consultation it developed that in order to meet system specifications this area -- one of the largest -- would have to be redesigned. All previous work would have to be scrapped. At the same time, it was decided that the services of approximately two thirds of the [REDACTED] of programmers STATINTL and analysts would no longer be required.

The second most troublesome problem facing the task force was the language barrier. This barrier existed between those responsible for ADP system design and everybody else. The people on the task force who encountered the most difficulty were the senior analysts from OCR. These people were of the "traditional library" world and though OCR

- 28 -

was one of the largest and earliest users of punched card equipment, the transition to digital computers was traumatic. The situation could hardly have been worse since this group of OCR analysts was responsible for interpreting requirements for the ADP system designers. The OCR people were experienced and skilled in establishing indexing procedures and languages or selection criteria for systems where the search is instituted by a knowledgeable human. However, they had difficulty in addressing the levels of specificity, completeness, and unambiguity which are mandatory if the search and storage mechanism is a computer program system.

A further level of confusion resulted from statements to the effect that the program system was to be a "powerful, flexible, generalized, and open-ended mechanism, capable of handling present and future data storage and retrieval requirement." How such a statement is interpreted depends, of course, on the sophistication and training of the listener. For example, "a serious accident" to a police officer may mean a collision of several vehicles and extensive bodily injury; to a high school student driving his first car, it may mean a damaged fender. Thus the OCR analyst came to expect^{the} CHIVE system would be a considerable improvement over the one that was currently in use in OCR; the ADP people felt that the CHIVE system would be very successful if it were able, initially, to do as well as the current system and in addition have growth potential. This difference of interpretation plagued the project throughout its life.

One can imagine the political repercussions attendant to a realization on the part of policy makers and top management that millions of dollars were being spent to develop a system whose basic design goal was to do as good a job with a computer based system as was currently being done by humans with punched cards and dossiers. Arguments, to the effect that the current system had reached its peak in efficiency and reliability and that further progress depended on adopting new techniques, were apt to sound specious or revolutionary.

The area of difficulty which ranked third in the hierarchy of headaches was the training problem. As noted previously, the CHIVE Task Force was charged to produce the tools and procedures required to make the China Operations Group a production unit. Subsumed under the phrase "tools and procedures" was the training necessary to acquire: area knowledge, subject matter expertise, familiarity with ADP processes, file definition capability, indexing competence, understanding of selection problems in a new environment, and so on. Much of this training could only come from osmosis since formal training programs in some of these technical areas did not exist.

Thanks to the considerable experience and capability of a few members of the OCR contingent, it was possible to develop training programs for indexers. A course consisting of on-the-job and classroom training was designed for the information analysts to give them area knowledge, document flow familiarity and so on. However, the course did not address the new techniques and language which the analyst must understand in order to use the computer

- 30 -

program system. Early in Phase III, the programming group chief had been asked to provide a series of seminars/lectures in order to enlighten the uninitiated on file management concepts and particularly on what the CHIVE file structure would be. This was not successful and was abandoned after it became clear that the communication problem successfully inhibited the learning process. Most of the OCR analysts attended ADP Orientation courses. These courses usually lasted five days or less and did not add much to the ADP sophistication or knowledgeability of the attendees.

On the other side of the house, i.e., the programming group, there was no effort to indoctrinate its members in the language or concepts of library science. This was to be expected in light of the severe demands of the program production schedule. If the program developers had to choose between designing programs and explaining the implications of the design, it was natural for them to select the former.

This rather bleak situation was relieved to a certain extent by one element of the project which marched to a different drummer, i.e., had a different schedule and a separate set of requirements. This small enterprise came to be known as SKAN and during its life was roundly reviled, staunchly supported, or greatly admired depending on the vantage point of the observer. The history of SKAN is somewhat checkered and will be reviewed briefly.

In mid 1965, a conversation between CHIVE analysts and OCR management started the SKAN design effort. The first requirement was to produce a KWIC (Keyword in Context) listing of all collateral

documents. All documents were to be indexed by title and area; documents of special significance would be subject indexed as well. The anticipated use of such a listing would be to answer questions on non-priority areas.

As often happens, requirements change. The new requirement was to produce a KWIC listing of current collateral receipts on Communist China which would be published periodically and disseminated to production and research analysts working on China problems. The burden of knowing how the periodical was to be read thus shifted from the OCR information analyst to the production or research analyst. This apparently minor change in requirements may very well have drastically diminished the usefulness of the SKAN product since production analysts don't mind reading documents but they wish to have the proper documents presented for their inspection with as little personal effort as possible. Dissemination of SKAN meant that they would have to read the listing to determine if they had missed any important documents in current receipts.

Other design changes were made as the analysts studied the problem. For instance, all documents selected were to be subject indexed - thus reducing, to a certain extent, the scope of collateral material included in the publication. Keyword in context was judged to be difficult to read so keyword out of context was chosen as the format. Initially conceived as an EAM application, the scope of the work and the availability of IBM programmers who

- 32 -

were too junior to contribute much to the system design of the large program system, led to a recommendation that SKAN be developed for the IBM 360/30. Thus SKAN became the bridge between computer programmers and some of the OCR analysts.

In its relatively short life, SKAN chalked up some impressive firsts. It was the first production program generated by CHIVE Task Force programmers. In fact, it may have been the first production program on an IBM 360/30 within the Agency. It was the first, and only, for some time, production use of the CDC-915 (an optical character reader system) in the Agency. This was a requirement for SKAN which did not effect the value of the product. Though initial data input for SKAN was on punched cards, it was decided that as soon as feasible, index records to the SKAN program should be input via the optical character reader. The purpose was to provide the designers and developers of the CHIVE system with data on the use of this new input device. The CHIVE Phase II Report had recommended the use of optical character reading for input transcription, but actual experience by Task Force personnel was very desirable. Therefore, when the CDC-915 Page Reader and associated program modifications were available, input was transferred to the Page Reader. This gave the CHIVE Task Force experience in both techniques on the same system, so comparisons and judgements could be made with some objectivity.

In addition to the virtues cited above, it also assisted in a number of other areas: forms design for an optical character reader, development of indexing codes, training of indexers, shakedown of the CDC-915, training of transcription typists, training of programmers

- 33 -

on the 8092 computer which was part of the CDC-915 system, and the production of programs for that computer. Finally, SKAN gave the programming group some insight into the problems associated with IBM's new line of hardware and software. However, the tremendous value of SKAN was not as an experimental device or even as a sop to CHIVE inquisitors, but rather as a mechanism to assist the education and orientation process. It was unfortunate that only about 15 people on the task force were actively involved with SKAN and so could benefit from the experience.

There was one other area within CHIVE that required programming services over and above those necessary to build the major system. This area was called Dictionary Building and in the CHIVE development schedule, dictionaries of geographic locations and installations (factories, research facilities, airfields, etc.) were of critical importance since they were needed to train the COG indexers and analysts.

At the time work began on building these dictionaries, no all-source dictionary existed on Chinese locations or installations. There were several dictionaries, all containing different data or similar data in different formats, in the Agency and the Community. It was, therefore, necessary to pull these dictionaries together and eliminate redundancies and contradictions in order to produce a CHIVE dictionary.

Since some of these dictionaries were available on magnetic tape, it was felt that the computer could be used to combine the dictionaries, do some rough editing, and provide those listings which experienced

- 34 -

analysts would require to do the final detailed editing.

The Formatted File System mentioned in Phase II and which was used in the Indexing Experiment looked like a good prospect to manipulate these dictionary files. An analyst/programmer trainee team was assigned to this area task. The reasons for choosing a trainee were: no experienced programmer was available, and apparently little or no programming was necessary -- it was simply a matter of manipulating FFS.

The difficulties encountered on this task caused schedule slippages which amounted to about 9 months. Fortunately the schedule had a certain amount of slack. Also other slippages masked the problems encountered on the dictionary task. A catalog of the software problems, machine problems, and file acquisition problems would be rather lengthy. Suffice to say that the three or four people involved in this task did build the files and learned a great deal in the process.

Fourth in the list of critical problems which faced the task force was the fact that it was a developmental endeavor. The difficulties which beset the Task Force are best characterized as managerial. These difficulties were of two major types: intrusion and confusion.

The confusion came about in a fairly straightforward way. It is best explained by elaborating on some of the idiosyncracies of developmental projects.

There are three major phases to acquiring a new system: design/

- 35 -

sales, development/engineering, installation/testing/production. If the system is acquired over a fairly long time period, the designers/salesmen are usually not around during the next phase, i.e., development/engineering. Similarly, system developers tend not^{to} be around for the production period though they might be available for the installation/test part of this phase. This metamorphosis is quite reasonable. Designers are an expensive, scarce commodity and cannot be squandered on a project once that phase is completed. Similarly, developers are scarce and expensive and must be moved to another project once their phase is completed. If there is little continuity insofar as people are concerned, the glue which keeps this development process from coming apart must be found elsewhere. And it is found in two words: requirements and specifications. The requirements for the system must be well-defined, complete, thoroughly understood, and capable of being translated into firm specifications. Upper management must play a major role in the requirements/specifications iteration in order that there be no surprises or recriminations downstream. The analogy which makes this point very well can be found in the dissatisfaction of someone who has engaged an architect/builder team to construct a house. If the owner of this new house complains of a deficiency, it is natural to ask if that item was a requirement. The answer comes back in many forms but usually can be put in one of three

- 36 -

categories: no -- it was not a requirement, yes -- it was a requirement but there was a breakdown in communication, or yes -- it was a requirement and the architect and/or builder is culpable.

CHIVE suffered from a lack of specific requirements. Lacking an appreciation of its critical role in establishing requirements, management approved the development of the CHIVE system primarily because it trusted the integrity and expertise of the designers/salesmen. Unable to obtain a complete set of requirements, the designers went forward with a design which was based on their understanding of the needs, attitudes, policies, future plans, and general milieu of the requesting office. (All of which, obviously, were subject to change without notice). The requirements/specifications iteration thus took place with very little contribution by upper management.

As soon as the designers/salesmen had completed their jobs (end of CHIVE Phase II), the development team started to come on board. These people began, or should have begun, asking searching questions about the lack of specifications. (As noted in the Phase II discussion, a system cannot be built or tested without specifications). The designers/salesmen were unable to obtain answers to some questions. Thus, the developers set their own specifications. The analogy of the dissatisfied homeowner is pertinent.

- 37 -

If the architect/builder team were to ask the customer for his requirements regarding the dimensions of a living room and the customer responded with: "adequate", the customer has thus renounced his responsibility and must live with whatever the architect/builder team considers adequate.

Once Phase III was nicely underway, the designers/salesmen phased out and OCR management -- the system operator -- was left with the developers/engineers. Three things impacted on the operator at this point. First, the team he trusted was no longer running the show and although the designers vouched for the developers, "things were not the same." Second, the developers were a different breed, with a different language and different capabilities than the designers. As a result, the carefully nurtured communication channels and devices which had been established by the designers and the operator were seriously disturbed. Third, the cost of the system took a sharp turn upward. Suddenly there were ninety people instead of nine, equipment was being ordered, space became a problem,

- 38 -

manpower allocation was torturous, and priorities for clerical and other support services had to be reassigned in recognition of the demands of the task force.

The natural result of these three perturbations was confusion. Confusion gave rise to intrusion. Upper management, not being quite sure of what was going on and under some pressure to justify rising costs and (later) schedule slippages, requested briefings, early delivery, cost effectiveness studies, system comparisons, and so on. In addition, the designers/salesmen had done a good job of making the system attractive to other offices, other directorates, and other members of the intelligence community. This interest resulted in more briefings and reports. The inevitable result of these examinations was the levying of new requirements or the alteration of old requirements.

The task force management did a good job of insulating the working level from these intrusions. However, it was inevitable that a large number of the senior people on the task force would become aware of what questions were being asked and the implication of some of the demands. The effect on morale can be appreciated.

Beginning in the summer of 1966 the excursions into the CHIVE Task Force for briefings, technical assistance, information, and writing ability became more and more frequent. One requester in OCR management wanted a compacted schedule for the completion of the tasks necessary to prepare the China Operations Group for operational responsibility. An in-depth study of OCR - its functions, products, responsibilities, and particularly the development effort embodied in CHIVE - was

- 39 -

initiated by the DD/I. Briefings took a considerable amount of time and CTF managers were also called upon to write a good deal. Indexing schemes which were less powerful than the CHIVE approach were considered and CTF analysts were asked to respond.

During FY67 COINS was underway and the CHIVE Task Force was expected to support this effort. In the same period OCR/BR began its development of the Soviet Elite File - later to have two names since the file had to be divided. OCR management called upon CHIVE for assistance and the special assistant to the D/CTF became the senior EDP analyst on this effort. Later the CTF supported this activity with other manpower contributions.

OCR Management recognized the reservoir of talent residing in the CHIVE Task Force in many ways. The preliminary work done by CTF personnel on a management system for COG encouraged OCR management to request these people to design a management system for all of OCR. The newly formed Information Processing Staff under OPPB has issued a call for all offices and directorates to prepare Five-Year ADP Plans - in effect ten years since the report had to cover the preceding years of ADP use as well as provide a five year plan for future ADP utilization. Members of CTF were tasked to coordinate, and contribute to, the OCR contribution as well as coordinate all office contributions for the DD/I's report.

Later, a team of analysts was directed by OCR to develop forms, gather data, conduct interviews - all to determine how OCR's services were used by Agency personnel, particularly within the Intelligence Directorate. CTF analysts were called upon to work on this effort.

- 40 -

As more and more of the CHIVE tasks were completed and the China Operations Group began staffing and planning, the unavailability of the computer program system (CAPRI) became a growing concern. Though coding on some parts of CAPRI had begun during the early weeks of the fall of 1966, other system components were still being specified. So it became quite clear that the target date of July 67 for initial COG operations using CAPRI was in grave danger of slipping.

This caused CTF personnel to consider alternative ways of testing some of the ^CHIVE system capabilities. A study was launched to determine what subsystems could be checked and by what means in the absence of a fully tested CAPRI. This resulted in the identification of some available equipment which could provide a rudimentary storage and retrieval capability. The proposal was approved in October 1966 and it was planned that the equipment would see use in February 1967. This never came to pass.

The depth of indexing which was basic to the CHIVE design came under the scrutiny of CTF personnel. This resulted in a proposal for a "fall-back" system which called for a different division of work for the indexer-analyst team and less indexing depth. Late in the spring of 1967 as CTF personnel were shifted to COG and training began, it was decided that an analyst-indexer team in COG would test the "fall-back" system.

Another activity which would impact on CHIVE later but which at the outset did not appear to warrant much attention by CTF personnel was the growing viability of the OCR computing facility. <

Early in FY67, OCR had launched a study to assess the feasibility

- 41 -

STATINTL

of substituting an IBM 360/30 for its extensive EAM holdings. The study was staffed with [REDACTED] and OCR personnel - CTF was not asked to participate. This study resulted in a recommendation to the DD/I that the computer be acquired. And it was - in late November 1966. Soon after the system was installed, OCR management approached the CHIVE Task Force to obtain programmers to support the conversion of EAM files and processes to EDP programs. Needless to say, the CAPRI developers were loath to slip their schedule any further by releasing programmers. Eventually two OCR trainee programmers in the Task Force were detailed to the Machine Division in OCR. This had an effect on activities such as the dictionary tasks and the Soviet Elite File effort.

During this same period the OCR computer was operating well but was not yet saturated with work. Consequently, the CAPRI programmers were able to get code check time on this computer. This helped considerably since the code check ^{and production} burden on the OCS computers was severe.

As a result of dedicated effort on the part of the programming group, it appeared likely, in the spring of 1967, that if not all, at least some CAPRI capabilities would be ready by July of 1967. Then the group managers discovered the design problem discussed earlier and the schedule for CAPRI slipped to a degree that OCR management considered unacceptable.

This coincided with the DD/I directive that OCR should reorganize along geographic lines - much earlier than CHIVE had planned; it happened at a time that OCR was beginning to have confidence in its

- 42 -

computer system; it occurred at about the time that OCR had provided a software contractor with the necessary guidance to provide an information storage and retrieval capability on the OCR computer; and it crystallized a decision that the depth indexing of the CHIVE system was too expensive for an office that had just suffered another reduction in force. The sum total of these forces had a predictable result: in July of 1967, OCR reorganized and the CHIVE Task Force was disbanded.

A list of major events during Phase III is as follows:

1. Oct 1965 All members of the CHIVE Task Force with the exception of the Associate Director and the programming group were colocated in 2G24. An all hands (CHIVE Task Force) meeting was held in the auditorium to discuss project objectives, career potential, organization and the like.
2. Nov 1965 The process of interviewing candidates for the OCR contingent of the Task Force began.
3. Dec 1965 The final report on conversion of Intelligence Reference Systems Index Records was disseminated.
4. Jan 1966 A working paper entitled Document Delivery System Organization and Space Recommendation was forwarded to OCR management. This completed the Task Force contribution to the upgrading of OCR's document recording and distribution capability.
5. Feb 1966 The chief of the OCR component of the task force (Development Staff) left the Agency and was replaced by a senior OCS analyst who transferred his career designation back to OCR.
6. May 1966 The Director/CHIVE Task Force left to become DD/OCS. The associate director of the Task Force (who was also chief of the OCR contingent) became D/CTF. The Task Force was reorganized to reflect the new assignment of responsibilities caused by this change as well as the resignation from the agency (effective in June) of the Chief/COG.

- 43 -

7. May 1966 The CDC 915 Page Reader System was delivered to OCS space in Headquarters building.
8. June 1966 The Security Procedures task personnel completed their major assignment, i.e., production of papers on physical security, personnel security, Top Secret controls, and the security augmentation problem.
9. Sept 1966 The Subject Keyword Announcement (SKAN) went into regular production and was disseminated weekly to about 25 analysts.
10. Sept 1966 The CDC 915 Page Reader System passed acceptance tests.
11. Sept 1966 The task team assigned to study the requirements of the China Operations Group for dossiers on ChiCom personalities completed the assignment. The working paper describing task team findings and recommendations was accepted.
12. Oct 1966 The indexers version of the ChiCom Location Dictionary was published.
13. Oct 1966 A chief for the China Operations Group was selected. He had served as an intelligence production officer in ~~PCI~~_O.
14. Nov 1966 A detailed study of the function, activities, and projects in OCR by a representative of the DD/I was completed. The CHIVE Task Force contributed substantially to the study and the final report.
15. Nov 1966 An IBM 360/30 computing system was installed in OCR space in Headquarters
16. Dec 1966 A report, entitled "SKAN Customer Survey Program" which summarized interviews with the recipients of the SKAN publication was issued. It recommended that the publishing of SKAN continue.
17. Dec 1966 A document, describing the input transactions for documents in COG, was published.
18. Jan 1967 A paper describing the proposed organizational structure for the COG was discussed and accepted.
19. Jan 1967 Copies of the ChiCom Location Dictionary were distributed within the Agency.

- 44 -

20. Jan 1967 A document describing processing procedures and standards for building ChiCom dossiers was published.
21. Jan 1967 The second edition of the last two volumes of the "Catalog of Document Series Reporting on Communist China" ^{was} published, thus completing this aspect of the Customer Requirements task.
22. Feb 1967 CHIVE Task Force managers ^{were} informed of ^{the} DD/I directive to D/OCR to reduce T/O level.
23. March 1967 The data input procedure for the SKAN process was changed from punched cards, to page reader forms input to the CDC 915 Page Reader System. Production of SKAN was uninterrupted.
24. May 1967 The chief of the Index and Retrieval Design Group, a senior OCS analyst, transferred to OCR and left the Task Force to work full time on the Area Reorganization Planning Group (ARPG) established by D/OCR. The ARPG was tasked to develop the plan, procedures, etc., for reorganizing OCR along geographic lines. The D/CTF was named as a member of the ARPG but was not relieved of his current responsibilities. The OCR instruction of 8 May announcing the ARPG concluded with the statement that, "The operational testing of Project CHIVE in the China Operations Group (COG) will continue as planned."
25. May 1967 Regular distribution of collateral and special intelligence materials to COG was initiated in order to train COG indexers and analysts.
26. May 1967 The Source Formatted Data task, which addressed the feasibility of taking data which came to the Agency in a formatted form and putting it into the computer with a minimum of manual processing, had progressed to the point of testing typewriters and other transcription devices.
27. May 1967 OCR entered into a contract arrangement with a software firm to modify ^{an} OCS program system to run on OCR's 360/30. The function of the modified program would be to provide information storage and retrieval.

28. May 1967 The unacceptability of the design of a major element of the CAPRI system resulted in a schedule slippage of at least 6 months.
29. June 1967 The modification of the Intelligence Subject Code, termed the Subject Indexing Code (SIC), complete with decode tables and a punched card deck to provide listings, was completed.
30. June 1967 The Management Information Task Team issued a paper setting forth a design concept for a management information system for the OCR CHIVE system.
31. July 1967 [REDACTED] on the CHIVE Task Force was reduced to five people, approximately one third its normal strength. Jurisdiction over the CAPRI effort was transferred to OCS. All OCS careerists were transferred to positions in OCS. STATINTL
32. July 1967 An OCR Instruction dated 7 July 1967 described the reorganization of OCR which resulted from ARPG deliberations. All elements of the OCR contingent ^{on the CHIVE Task Force} were disbanded except for COG which would continue testing ^{as the Experimental Facility} under the direction of the newly formed Planning Management Staff. The SKAN Indexing Unit was placed under the direction of Chief/~~OCF~~ *Experimental Facility*.